

IB9XC-15 Governance, Risk and Behaviours in AI-augmented Systems

27/28

Department

Warwick Business School

Level

Taught Postgraduate Level

Module leader

Michael Mortenson

Credit value

15

Module duration

9 weeks

Assessment

100% coursework

Study location

University of Warwick main campus, Coventry

Description

Introductory description

As organisations transition from passive software tools to autonomous AI agents capable of independent corporate action, traditional governance models are no longer sufficient. This shared module bridges the gap between Information Systems and Behavioural Science to address how managers can design robust oversight frameworks.

Managing an agentic enterprise requires a dual approach: implementing technical safeguards to constrain digital behaviour and understanding the psychological shifts that happen when human teams delegate decisions to AI. Students will explore how to architect compliance frameworks that integrate programmatic guardrails with real-world human oversight, ensuring safety without stalling operational innovation.

Module aims

This module incorporates a balanced range of oversight, compliance, and behavioural topics relevant to governing Agentic AI developments in modern enterprises, including

Analysing the socio-technical risks that emerge when autonomous systems interact with human

organisational structures.

Understanding trust calibration - how to prevent human over-reliance (automation bias) or irrational rejection (algorithmic aversion) of digital agents.

Designing human-in-the-loop and human-on-the-loop oversight workflows that account for cognitive limits and alert fatigue.

Developing corporate accountability frameworks, audit trails, and legal liability boundaries for autonomous system actions.

Translating global AI regulations and compliance mandates into specific system constraints and corporate operating procedures.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

The module will cover the following topics:

Socio-Technical Governance Frameworks: Integrating corporate policy with technical systems architecture.

Trust Calibration & Cognitive Biases: Managing workplace pathologies like automation bias (blind compliance) and algorithmic aversion (rejection of the technology).

The Agentic Risk Landscape: Identifying novel operational liabilities, including hallucinations, data drift, and unauthorized execution chains.

Programmatic Guardrails: Conceptual design of real-time technical constraints, boundary definitions, and input/output filters.

Designing Human-on-the-Loop Oversight: Engineering human intervention checkpoints that match human cognitive capacity and prevent passive rubber-stamping.

Accountability, Liability, and Corporate Policy: Defining legal, financial, and organizational ownership when an autonomous workflow encounters errors.

Regulatory Compliance: Navigating major global regulatory frameworks and translating them into corporate policy.

Auditability and Forensic Logging: Designing system telemetry and tracking metrics to support post-incident investigations.

Incident Response & Strategic Case Studies: Crisis management protocols and organizational recovery when automated agent workflows fail.

Learning outcomes

By the end of the module, students should be able to:

- Demonstrate a sophisticated and practical understanding of Agentic AI concepts, use cases, and risks.
- Critically evaluate Agentic AI use-cases and applications with key consideration of ethical, strategic, and operational dimensions.
- Evaluate legal and regulatory requirements to translate abstract compliance mandates into concrete operational boundaries.

Indicative reading list

[Reading lists can be found in Talis](#)

[Specific reading list for the module](#)

Subject specific skills

Deploy agent system design, implementation, and monitoring with an explicit focus on risk mitigation.

Design comprehensive corporate risk-assessment matrices and operational guardrails for autonomous systems.

Configure human-agent workflow checkpoints that optimize supervisor attention and minimise cognitive fatigue.

Formulate clear, compliance-ready corporate AI policies that satisfy legal standards while ensuring organisational adoption.

Formulate comprehensive socio-technical governance blueprints that combine digital safeguards with realistic human oversight loops.

Transferable skills

Balanced systems thinking, integrating technical constraints with human behaviour.

Interdisciplinary collaboration between technical (IS) and managerial/HR functions.

Risk assessment and strategic decision-making under technological uncertainty.

Written defence of compliance, risk, and corporate governance strategies.

Study

Study time

Type	Required
Lectures	9 sessions of 1 hour (6%)
Other activity	18 hours (12%)
Private study	49 hours (33%)
Assessment	74 hours (49%)
Total	150 hours

Private study description

Private study to include preparation for lectures/ seminars/ workshops and own reading

Other activity description

9 x 2 hr workshops F2F

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Assessment group A

	Weighting	Study time	Eligible for self-certification
Assessment component			
Individual Assignment	80%	59 hours	Yes (extension)
Individual Assignment 2500 words			

Reassessment component is the same

Assessment component			
Group project and video presentation	20%	15 hours	No

Reassessment component

Individual Assignment	Yes (extension)
-----------------------	-----------------

Feedback on assessment

via my.wbs

Availability

There is currently no information about the courses for which this module is core or optional.